76 327.1

REGIONAL TRANSPORTATION AUTHORITY

METHODS OF PREDICTING LOCAL DEMAND FOR NEW RAPID TRANSIT SERVICES

by

Inwon Lee

&

Howard Permut

TR-75-03

TRANSPORTATION LIBRARY
NORTHWESTERN UNIVERSITY LIBRARY

TECHNICAL REPORT

TRAN
HE
318
L478

CHICAGO, ILLINOIS

3 5556 031 83118

METHODS OF PREDICTING LOCAL DEMAND FOR NEW RAPID TRANSIT SERVICES

by

Inwon Lee

&

Howard Permut

TR-75-03

Technical Report Series Presented By

Planning and Development Department
Regional Transportation Authority
(Northeastern Illinois)
P. O. Box 3858
Chicago, Illinois 60654

December, 1975

TRAN
HE
318
L478

OVERVIEW

This paper summarizes two models developed by the Illinois Department of Transportation[1] to predict local travel demand, as well as four alternative approaches developed by the Chicago Transit Authority. The concepts behind the CTA work are similar to those used by IDOT except that slightly different analytic methods and a more detailed data set were used.

Both the CTA and IDOT analyses emphasize empirical travel patterns and are formulated so as to predict future levels of ridership on new rapid transit lines based on actual demand levels on existing rapid transit facilities.

IDOT ANALYSIS APPROACH

Introduction

The IDOT analysis was undertaken in order to estimate potential ridership on a crosstown rapid transit line. The models developed assume that crosstown ridership would be comprised of local non-CBD trips. Since, there is no reason to believe that local radial and local crosstown trips would be behaviorally different, the IDOT models can be used to predict local ridership on radial as well as crosstown routes.

IDOT Method One: Local Ridership as a Function of Employment and Population

A linear regression model was constructed to predict local ridership on a rapid transit line as a function of the population and employment densities of the line's service area. The service area was defined as one mile on either side of the line.

As noted above, the underlying assumption was that the proposed crosstown line would be similar to existing rapid transit lines in all respects except that it would not serve the CBD. Thus the relationship between its local ridership generation and employment/population densities would be the same as that of existing lines.

---

1 Burns, I., "Estimated Locally Generated Rapid Transit Demand in the Crosstown Corridor," May 1974.
Hovind, M. and Lisco, T., "Estimated Potential Ridership for Alternative Proposed Crosstown Public Transit Systems", June 1973

This method was based on the effect of the Dan Ryan Line on neighboring South Side bus lines. The entering volumes at Dan Ryan Stations as well as ridership, before and after the Dan Ryan was opened, on neighboring South Side bus lines was gathered. Two adjustments of the data were made to account for: (1) riders, south of 63rd Street, who had previously used the parallel bus routes as feeders to the Jackson Park/Englewood Branches but now use east-west bus routes as feeders to the Dan Ryan; and (2) riders who still used the parallel bus routes as feeders to the Jackson Park/Englewood Branches.

The following line was then fit through the data points for seven paralleling routes:

$Y = .46 - .14X$

where:

Y: percentage diversion from bus route
X: distance (in miles) between bus route and rail line

The one problem with this methodology is that no statistical relationship between percent diversion and distance from the rapid transit line was obtained.

Summary of IDOT Methods

The IDOT models seemed to be addressing the correct variables, but with little success. Four alternative approaches, two of which are regression analyses based on the first IDOT method, and two of which are factors based on empirical evidence were then developed.

CTA ANALYSIS

Introduction

The CTA methods were formulated to address the following points:

1) Develop local demand models which can be easily used in an on-going rail rapid transit system optimization study.

TABLE 1

Ridership, Population and Employment Statistics of
Chicago Transit Authority Rapid Transit System

Line	Route Length From Terminal to CBD Boundary[1] (in miles)	Total[a] Population	Total[a] Employment	Population Density[2]	Employment Density[2]	Number of[b] Local Trips	%	Number of[b] CBD Trips[3]	%	Number of[b] Total Trips[4]
Howard & Ravenswood	14.7	406,573	146,838	27,658	9989	10,070	.126	52,035	.650	79,995
Congress & Douglas	15.1	320,528	138,331	21,227	9161	8600	.207	21,645	.521	41,565
Englewood & Jackson Park	10.7	220,855	75,789	20,650	7084	5645	.139	17,765	.438	40,515
Dan Ryan	10.8	175,382	76,972	16,240	7127	5140	.170	19,905	.657	30,290
Milwaukee	8.1	166,519	63,059	20,558	7785	4725	.136	22,920	.657	34,865
Lake	8.0	121,064	90,408	15,133	11,301	2760	.107	19,840	.768	25,840
Total	67.4	1,411,031	591,407	121,466	52,447	36,940	.146	154,110	.609	253,070

(a) 1975 Population & Employment (It is assumed that 1975 data accurately represent 1970 data). Projection of the Northeastern Illinois Planning Commission

(b) CATS 1970 Home Interview Survey

1 The CBD is defined as the area bounded by Chicago/Halsted/Roosevelt. For branch lines, the distance is measured from the terminals to the junction, and from the junction to the CBD.

2 Densities are calculated within one mile on either side of the transit line

3 This is the number of round trips with origins on the line and destinations in the CBD

4 Total trips are not the sum of local and CBD trips, since they include thru trips which originate on a line, pass thru the CBD and have destinations on another line

5) Determine the regional quotient,[1] and apply it in conjunction with the employment/population ratio. The regional quotient would distinguish between the service areas of lines with high and low concentrations of employment; and would presumably produce greater percentages of local trips for lines serving areas with larger employment concentrations.

The first method to be described is similar to the second regression equation developed by IDOT except that the number of local trips is specific to each line and a system-wide average is not used. The second method is also a regression analysis which expresses the percentage of local trips as a function of the employment/population ratio.

The other two methods utilize trip generation rates based on the existing system. The third method is based on the local trips generation rate for existing CTA lines. The fourth method is similar to the third except that the trip generation rate is modified by the regional quotient.

Once the models are developed, a comparison of their accuracy is undertaken. Then the use of the models is discussed.

CTA Method One: Local Ridership as a Function of Population Densities

Based on the data in Figure 1, a simple regression equation which describes the number of local trips per mile on a line as a function of the population density of the line's service area was obtained:

$$Y = \underset{(97)}{44} + \underset{(.005)}{.024}X_1 \qquad R^2 = .87$$

where:

Y: number of local trips per mile

X_1: service area population density (per two square miles)

---

[1] The functional form of the regional quotient is:

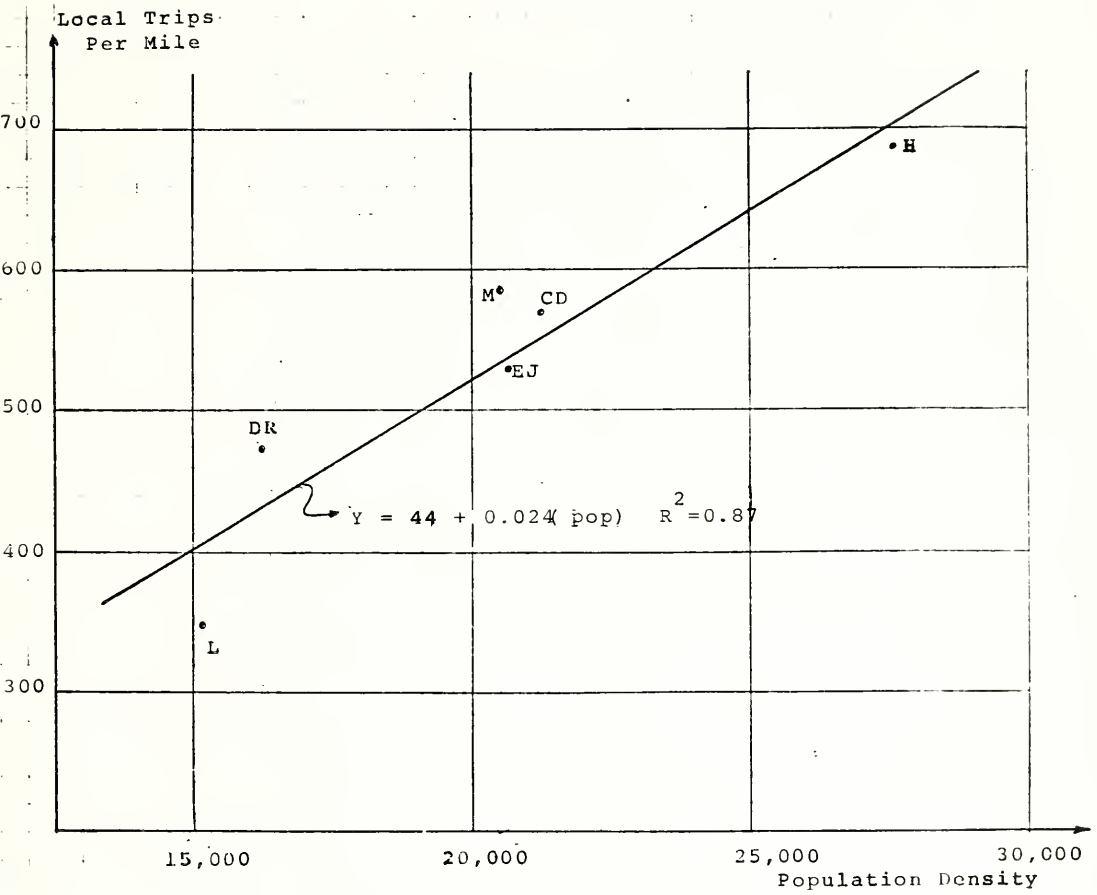
$$Q_x = \frac{\left(e_x / p_x\right)}{\left(\sum_{x=1}^{n} e_x\right) \Big/ \left(\sum_{x=1}^{n} p_x\right)}$$

where:

e_x: service area employment density of line x

p_x: service area population density of line x

FIGURE 1


Local Trips
Per Mile
700
600
500
400
300
H
M
CD
EJ
DR
L
Y = 44 + 0.024(pop) R2=0.87
15,000
20,000
25,000
30,000
Population Density

This equation is based on the data in Figure 2. The equation obtained was:

$$Y = -.097 + .645X \qquad R^2 = .69$$
$$(.098) \quad (.249)$$

where:

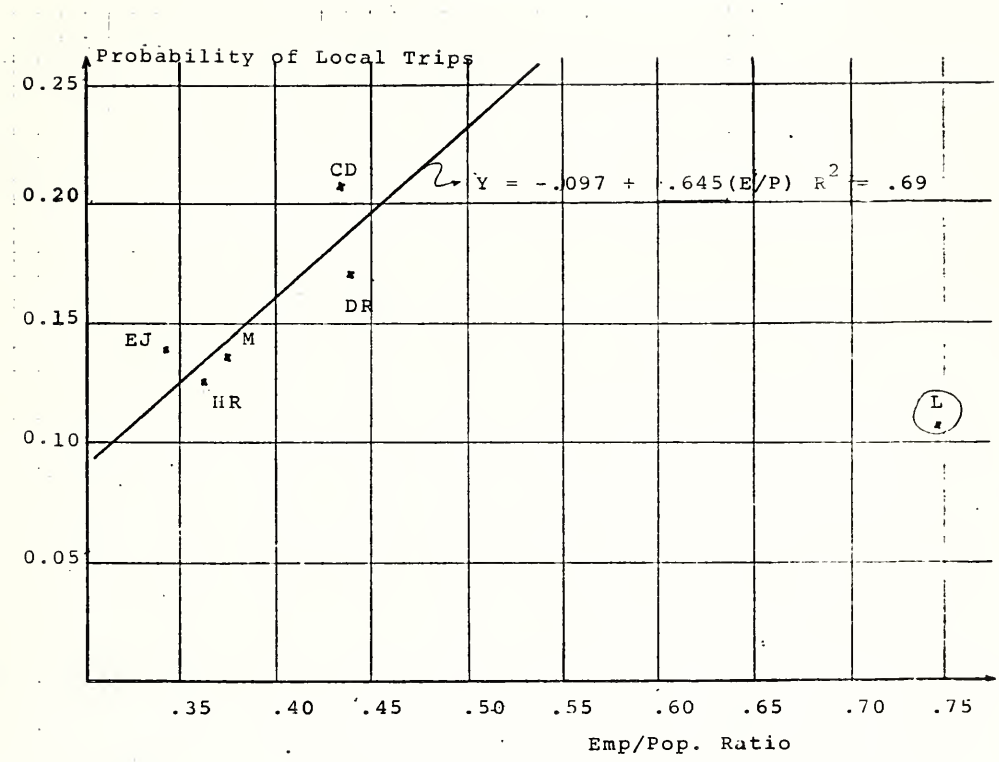
Y: percentage of local trips per line
X: employment/population ratio

This equation does not include the Lake Street Line since it is a significantly different data point, as is illustrated in Figure 2. The Lake Street line has both a low local trip generation rate and a high employment/population ratio, with this high per capita employment being the main reason for the unusualness of the Lake Street Line.[1] The low local trip generation rate is due to the fact that the line serves two areas with distinctly different socio-economic characteristics (Chicago and Oak Park) which have few trips between them, and that few trips are made within each area (Oak Park is too small for many local trips to use rapid transit, and very few people have origins or destinations on the Chicago stations of the line). The high employment/population ratio is due to a large number of manufacturing jobs (i.e. factories) and a low population density in the line's service area.

The results of this analysis are somewhat promising in that the hypothesized relationship does exist (i.e. the coefficient of the employment/population variable is positive). However, the relationship is not statistically significant at the 95 percent level ($6.7 < 10.1$, where $F_{1,3,.05} = 10.1$). Furthermore, neither the coefficient of the population variable nor the constant term are significantly different than zero at the 95 percent level ($2.6 < 2.8$ and $.98 < 2.8$ respectively, where $t_{4,.95} = 2.8$).

[1] It should be noted that, as shown in Methods One and Three, local trips on the Lake Street Line exhibit the same functional relationship with population as the other rapid transit lines.

FIGURE 2


Probability of Local Trips
0.25
0.20
0.15
0.10
0.05
CD
Y = -.097 + .645(E/P) R2 = .69
DR
EJ
M
HR
L
.35
.40
.45
.50
.55
.60
.65
.70
.75
Emp/Pop. Ratio

Furthermore, in comparison to the first regression equation, this equation does not have as a high correlation. However, this may be due to the fact that is it based on ratios[1] as opposed to total numbers.

However, this method is somewhat more causal in nature in that, unlike method one, it provides estimates of local ridership which are sensitive to changes in causal parameters (i.e. employment within the line's service area).

A second equation was developed to test if the hypothetical relationship between the service population ratio held for CBD trips (see Appendix B). This hypothesizes that the greater the ratio, the lower the percentage of CBD trips. However, this hypothesis was not borne out, and the relationship between the percentage of CBD trips and the employment/population ratio was statistically insignificant.

CTA Method Three: Local Ridership as a Function of Trip Generation Rates

Since, in Method One it was found that the constant term in the regression equation is not significantly different than zero at the 95 percent confidence level, local ridership can be accurately expressed in terms of a trip generation rate. For this reason, a simple relationship which expresses the number of local trips on a line as a function of the population of the line's service area as well as a regional trip generation rate was obtained:

$$LT_x = (P_x)\ (\frac{LT_T}{P_T})$$

where:

LT_x: number of local trips on line x

LT_T: number of local trips on all rapid transit lines

P_x: population of line x's service area

P_T: total population within service areas of all rapid transit lines.

LT_T/P_T: local trip generation rate

[1] Regression equations based on rates or ratios usually have lower correlations because the variance of both the dependent and independent is smaller, and the entire variance is due only to rate variance and not to service area size variances.

The trip generation of the CTA lines is .026 local trips per person.

The results of applying this trip generation method to all CTA lines are shown in Table 2.

TABLE 2

Estimation of Local Trips by the Trip Generation Rate

Line	Estimated Local Trips	Actual Local Trips	Error	% Error
Howard & Ravenswood	10,644	10,070	-574	5.7
Congress & Douglas	8,391	8,600	209	2.4
Englewood & Jackson Park	5,784	5,645	-139	2.5
Dan Ryan	4,592	5,140	548	10.7
Milwaukee	4,359	4,725	366	7.5
Lake	3,169	2,760	-409	14.8

The correlation between the observed and estimated number of trips is extremely high ($r^2 = .97$).[1] Thus, this method, while having no underlying causal theory, provides very accurate estimates of local demand for existing rapid transit lines.

A test relating CBD travel on a line to its population and a CBD trip generation rate was undertaken (see Apendix C). As was found for the first method, the errors in the estimated CBD ridership (as great as

[1] It should be realized that this correlation value cannot be directly compared to the correlation coefficients developed for Methods 1 and 2, since the independent and dependent variables (as well as the measurement units) for each method are dissimilar.

62 percent) were far greater than the errors for the estimated local ridership. The reason for this are the same as those noted for method one, i.e. problems with service area definition.

CTA Method Four: Local Ridership as a Function of a Modified Trip Generation Rate

This is similar to method three except it has slightly more underlying causality in that the local trip generation rate is modified by a measure of employment, the regional quotient. The regional quotient is defined so that the probability of a local trip on a given line is proportional to both its employment/population density ratio and to the ratio of its employment/population density ratio and the average regional employment/population density ratio.

The functional form of this method is:

$$Y_x = (P_x) \times \left[\frac{(LT_t)}{(P_t)} \right] \times Q_x$$

where:

Y_x: number of local trips on line x

P_x: population within service area of line x

P_t: total population within service areas of five selected rapid transit lines

LT_t: number of local trips on five selected rapid transit lines

Q_x: regional quotient of line x

This method was applied to all existing rapid transit lines except the Lake Street Line, since, as shown in Method Two, the employment characteristics of the service area of the Lake Street Line exhibit a significantly different relationship with local trip generation than is found for the other lines.

The results of applying this method to five Chicago lines are shown in Table 3.

TABLE 3

Estimation of Local Trips by Modified Trip Generation Rate

Line	Regional Quotient	Estimated Local Trips	Actual Local Trips	Error	% Error
Howard & Ravenswood	.9333	9934	10070	136	1.4
Congress & Douglas	1.1152	9358	8600	-758	8.8
Englewood & Jackson Park	.8866	5128	5645	517	9.2
Dan Ryan	1.1341	5207	5140	-67	1.3
Milwaukee	.9785	4265	4725	460	9.8

The correlation between the observed and estimated number of trips is extremely high ($r^2 = .97$). This is similar to the results of Method 3, which should be expected since the modifications due to the regional quotient are relatively small.

Comparison of Methods

A comparison of the degree to which each alternative method can replicate existing data is shown in Table 4. Three measures of accuracy are shown: average percentage error, root mean square error, and percent root mean square error.[1] The absolute value of these measures for a specific method is not as important as the relative values for the different methods.

[1] The functional form of each measure is as follows:

average percentage error: $$\sum_{x=1}^{N} \left(\frac{|O_x - E_x|}{O_x} \right) \times \frac{100}{N}$$

root mean square error: $$\sqrt{\sum_{x=1}^{N} (O_x - E_x)^2 / N}$$

percent root mean square error: $$\sqrt{\sum_{x=1}^{N} (O_x - E_x)^2 / N} \times \frac{100}{\overline{M}}$$

where:

O_x: the observed (actual) number of local trips on line x

E_x: the expected (predicted) number of local trips on line x

$\overline{M}$: the mean value of the observed number of local trips for all lines

N: the number of lines included in each analysis methods (For Methods 1 and 3, this includes all rapid transit lines (N=6), while for Methods 2 and 4, the Lake Street Line is eliminated (N=5).)

TABLE 4

Comparison of Errors

Method	Average Percent Error	Root Mean Square Error	Percent Root Mean Square Error
Method One	6.8%	334	5.42%
Method Two	9.9%	716	10.47%
Method Three	7.3%	407	6.61%
Method Four	6.1%	464	6.79%

Table 4 shows that all four methods provide accurate results and that Methods One, Three and Four provide relatively equivalent results. For the root mean square error and the percentage root mean square error, the methods which are based solely on population (i.e. have less underlying causality) are superior to the ones which include both employment and population. Of the two population-based methods, Method One is superior. However, for the average percent error, Method Four provides slightly better results than Method One.

Summary of CTA Analysis and Recommendations For Application

The four alternative approaches improve on the IDOT work. The relatively accurate results obtained by all four methods are encouraging and suggest that one or more of these methods is applicable for the estimation of local demand on new lines in Chicago. For new suburban rapid transit lines, the models should provide accurate results, though, prior to their use, they should be validated using existing suburban data (such as Evanston).

The models should be used with great caution in estimating local demand on proposed new rapid transit lines in other cities. The application of the models is most valid for cities with similar local rapid transit trip generation rates and similar rapid transit local trip service areas. In more general terms, this implies that the bus is the primary mode for local trips, and that there is bus service every half mile. It is recommended that prior to application, the accuracy of the models be determined for the specific area.

Based on the comparison of the various methods, it is recommended that Method One, the expression of local trip generation as a function of population density, be used. If for theoretical reasons the inclusion of the impact of employment on local trip generation is desired, Method Four, the local trip generation rate modified by the regional quotient

APPENDIX A

CBD RIDERSHIP AS A FUNCTION OF POPULATION DENSITY

An analysis of the relationship between CBD trips on a rapid transit line and the population density of the line's service area was undertaken to enable a comparison with the local trip/population relationship previously obtained.

While the specific trip rate (i. e. slope of the regression line) was expected to be substantially different, the nature of the relationship (direction of the regression line) and the degree of correlation was expected to be similar.

Figure 3 shows the relationship between CBD trip rates and population densities. The following regression equation was obtained from this data:

$$Y = 413 + .093x$$
$$(1602)\ (.078) \qquad r^2 = .27$$

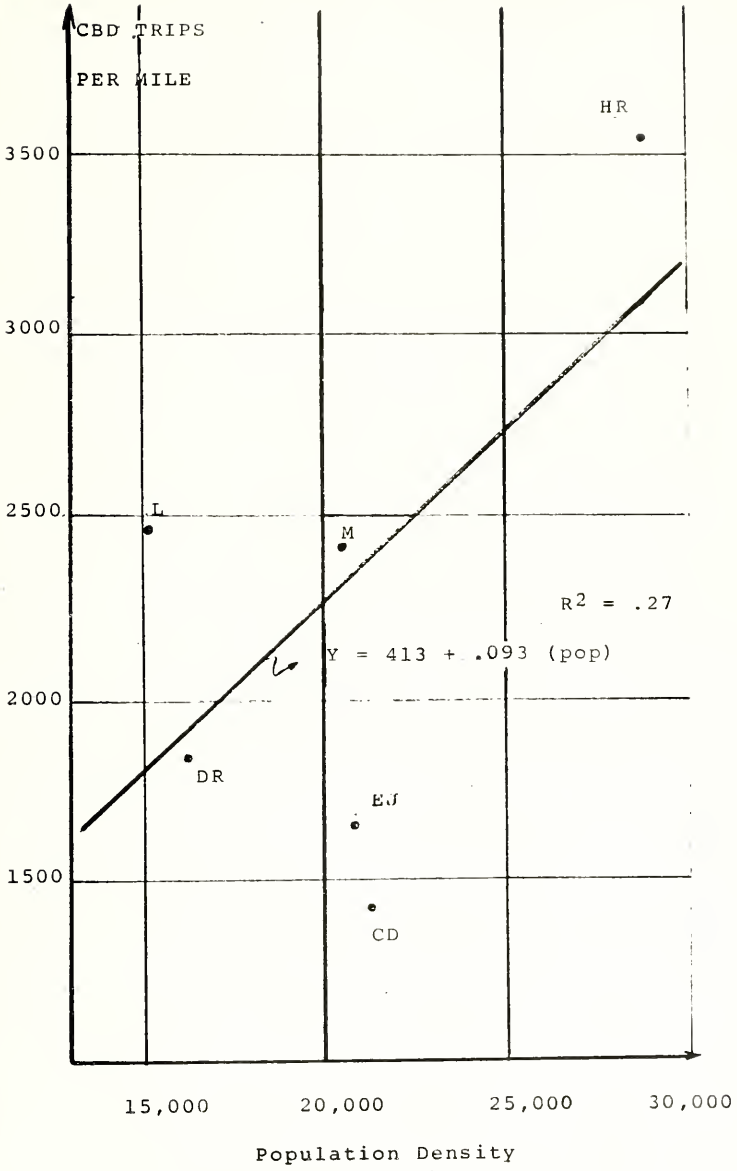
where:

Y: number of CBD trips per mile
X: service area population density (per two square miles)

As Figure 3 shows, the relationship between CBD trips per mile and service area population density is as hypothesized (the greater the population density, the greater the CBD trip rate). Also, the slope of the regression line is four times greater than was obtained for local trips. This is as expected since the rapid transit trip generation rate is much greater for CBD trips than for local trips.

However, as the regression equation illustrates, the degree of correlation between CBD trip rates and population densities is much lower than for local trip rates. The relationship between CBD trip rates and service area population densities is not statistically significant at the 95 percent level ($1.4 < 7.7$, where $F_{1,4,.05} = 7.7$). The reasons hypothesized for this are: (1) a one mile service area includes almost all the area from which local, but not CBD, trips originate (since for local trips, buses can provide a competitive service within a fairly small distance of a rapid transit line); and relatedly, (2) the distance from which people travel to a rapid transit line for a CBD trip varies greatly between lines

FIGURE 3


CBD TRIPS
PER MILE
3500
3000
2500
2000
1500
HR
L
M
DR
ES
CD
R^2 = .27
Y = 413 + .093 (pop)
15,000
20,000
25,000
30,000
Population Density

dependent on the nature of the line (a feeder-bus or a walk-in line).

In summary, the population density of a line's service area should not be used to estimate CBD trip rates on the line. However, this has no impact upon the validity of using service area population densities to predict local trip rates on a line.

APPENDIX B

PROPORTION OF CBD RIDERSHIP AS A FUNCTION OF EMPLOYMENT/POPULATION RATIO

An analysis of CBD ridership as a function of the employment/population ratio was undertaken to test the validity of the underlying hypothesis of the second method: the greater the employment/population ratio, the lower the percentage of CBD trips. Also, a comparison with the local trip employment/population ratio relationship was desired.

Figure 4 shows the relationship been the proportion of CBD ridership and the employment/population ratio. The following regression equation was obtained from this data, without considering the Lake Street Line:

$$Y = .33 + .646x \qquad r^2 = .07$$
$$(.52)\quad (1.3)$$

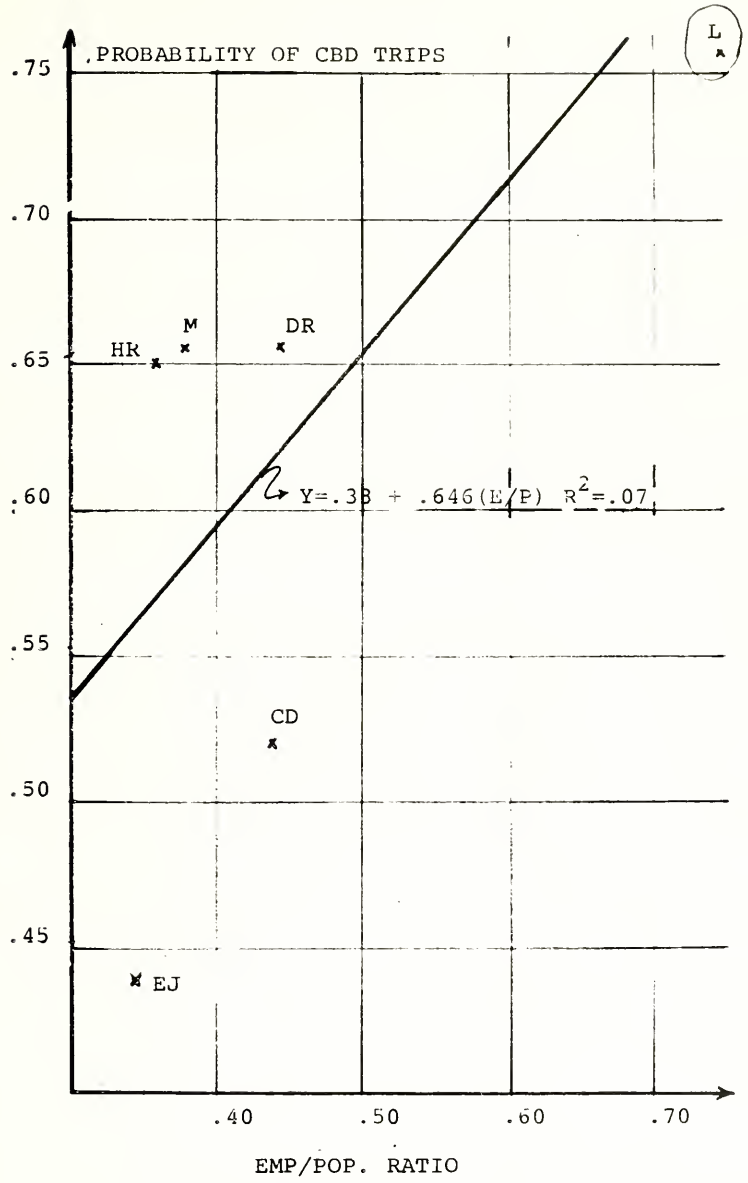
where:

Y: proportion of CBD trips
X: employment/population ratio

As the regression equation illustrates the underlying hypothesized causality does not exist. Furthermore, the relationship between CBD trips and the employment/population ratio is totally insignificant ($.24 < 10.13$, where $F_{1,3,.05} = 10.13$).

Due to the statistical insignificance of this relationship, it is felt that no conclusions concerning the underlying validity of the use of the employment/population ratio can be reached.

FIGURE 4


PROBABILITY OF CBD TRIPS
.75
.70
.65
.60
.55
.50
.45
L
M
DR
HR
Y=.38 + .646(E/P) R^2=.07
CD
EJ
.40
.50
.60
.70
EMP/POP. RATIO

APPENDIX C

CBD RIDERSHIP AS A FUNCTION OF TRIP GENERATION RATES

This analysis compares the accuracy of using trip generation rates to predict CBD trips as opposed to using trip generation rates to predict local trips. This is similar to the analysis done for Method One in Appendix A in that the accuracy of the method, as opposed to an underlying causal hypothesis, is being evaluated.

The functional form of the relationship is the same as Method 3 except that CBD trips are estimated. The trip generation rate for CBD trips is based on all rapid transit lines, and is .10.

The results of applying this rate to the existing rapid transit lines are shown in Table 5.

TABLE 5

Estimation of CBD Trips by the Trip Generation Rate

Line	Estimated CBD Trips	Actual CBD Trips	Error	% Error
Howard & Ravenswood	44405	52035	7630	14.7
Congress & Douglas	35008	21645	-13363	61.7
Englewood & Jackson Park	24132	17765	-6367	35.8
Dan Ryan	19156	19905	749	3.8
Milwaukee	18187	22920	4733	20.7
Lake	13223	19840	6617	33.3

The correlation between the observed and expected values is not high ($r^2 = .61$). The percentage errors are also fairly high, with an average percentage error of 28%.

In summary, the estimation of CBD trips in terms of a trip generation rate does not provide accurate results. However, this has no impact upon the validity of predicting local trips in terms of a trip generation rate.